Data Analysis for Investments

Course: Applied Statistical Methods (MATH -F432)

I/C : Dr.Sumanta Pasari

By Group 6

RAO GROUP

1	**SP Prusty**	**2023A2PS0235P**
2	Rudra Narayan Bhatt	2023A2PS0296P
3	Mohit Bugalia	2023D2TS1247P
4	Ansh Tayal	2023A1PS0189P
5	Harsh Khetan	2022A4PS0719P
6	Aryan Singh Sirohi	2023A1PS0173P
7	Gaurav Kataria	2022B4PS1525P
8	Tavish Singh	2023A8PS0718P
9	Ayush Aman	2023A4PS0228P

Table of Contents

1. NSE - National Stock Exchange

Established in 1992, NSE is the biggest Stock Exchange in India. It provides a platform for online trading. It is also helpful in depicting the patterns of Economy, the various trends in sectors, and provides a basis of currency exchange in India. Before its establishment, the trading was done manually but post that, it was made completely electronic which made the whole process more efficient, faster, accurate and transparent. It is regulated by SEBI- Securities and Exchange Board of India. NSE uses two major indexes - Nifty and Sensex.

2.1 Nifty 50 and its Methodology

Nifty 50 is the benchmark index of NSE. It takes into account the top 50 companies which have high capitalization (together they make more than 60% of the total free-float market capitalization of the country hence considered most efficient), high liquidity, that is it is easier to buy or sell the stocks of that particular company and are actively traded. Nifty 50 shows the overall performance of the Indian stock market. The methodology used in Nifty 50 is **Free-Float Market Capitalization**. Here the free-float market capitalization (market value) of each company is calculated as: Price of stock X Number of total shares available for public trading for that stock. This ensures fairness in trading as the shares which are not available for trading have no impact on the market value and is thus not influenced by promoters. Then a weightage(index) is assigned to the company as: Free-Float Market Capitalization of that company/ Total Free-Float Market Capitalization. For example, if a company has 10% of total Free-Float Market Capitalization then it gets 10% weight in the index. Depending on the weightage, each company is assigned some Nifty points from the total and the company is then listed accordingly in the NSE.The total Nifty points are calculated as the current Total Free-Float Market Capitalization / Total Free-Float Market Capitalization in the Basis year that the year when NSE became fully Operational (1995). The base index value is 1000. The listing is normally done twice a year (March and September) however, a company can be added or removed from the list during certain cases such as bankruptcy. It is also easily affected by global events. The Nifty 50 method also covers all the various sectors such as IT, Finance, Pharmacy etc. hence it is considered to be the most efficient in current times.

2.Why Do We Need to Invest?

Introduction

Investing is one of the most fundamental financial activities that help investor to grow their wealth and protect themselves against economic uncertainties. Although many individual deposits their earnings in a bank to get nominal interest return, it may not be sufficient in the long term. This report examines the various aspect of investing like why it is necessary, its psychological impact and discuss why simple bank deposits may not be sufficient in the modern world.

2.1 The Economic Rationale for Investment

The core principle of investment is that money is a resource that is capable of multiplying itself when deployed effectively and could lose value if store passively. In investment we use various financial resources - such as equities, bonds and mutual funds with core motivations such as:

- Capital Appreciation
- New income generation
- Inflation Protection

- Retirement planning

2.2 Wealth Creation vs. Bank Savings

The table below shows a comparison on different annual return based on historical returns

Asset Type	Approximate Historical Average Return (India)
Savings Account	**~2.5%–4% p.a.**
Fixed Deposits / Time Deposits	~5%–8% p.a.
Equity Market (Large-cap / Broad market)	~10%–13% p.a
Hybrid Funds / Equity-Debt Balanced Funds	**~11%–16% p.a.**

While Bank deposit may have an edge for safety and liquidity; the return on investments are clearly multiple times of bank deposits

2.3 Inflation vs. Bank Savings

Inflation is one of the most important factor often overlooked when putting money in bank accounts, if Inflation rate is higher than annual return provided by the bank, it effectively means that your money purchasing power is reducing rather than increasing, if we look at data from the year 2012-2021, we will see that at least for 3-4 years inflation is higher than 3 months yield , making a case to invest money in different financial tools.

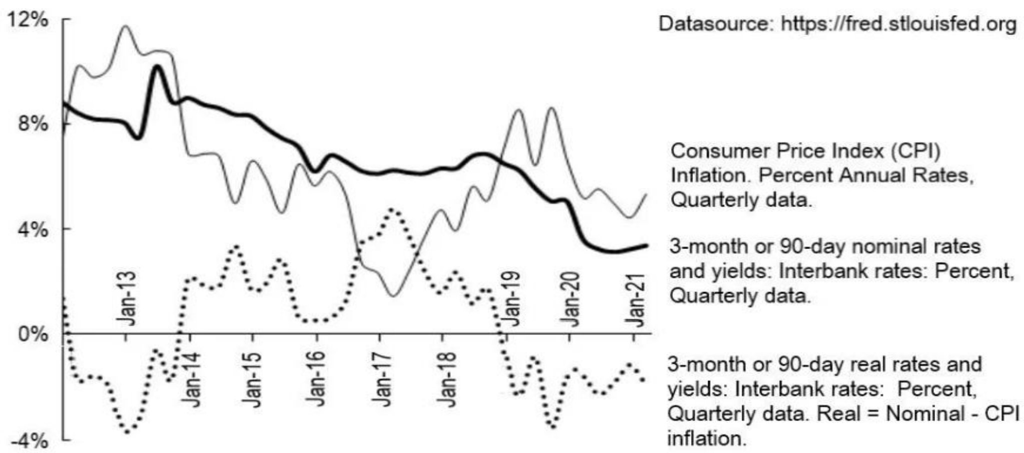


Fig 1: Consumer Price Index inflation from 2013 to 2021

2.4 Tax Benefits of Investing

Another key factor that makes investing more appealing are the tax saving schemes that are often associated with investing, some of the schemes that are famous and are easy to use are:

- **ELSS (Equity-Linked Savings Schemes**): In this Scheme, individuals can have Deductions up to ₹1.5 lakh per financial year, which reduces our total taxable income.

- **Public Provident Fund (PPF):** This falls under the “Exempt-Exempt Exempt” category which ensure that money used in investment and gained from Interest, Withdrawal are fully tax free

2.5 Behavioural Finance Perspective: Stress vs. Strategy

Often many people feel that investing money increases stress due to market volatility and also due to better liquidity option in banking. This is partially true only for short term investing. However:

- Long-term investing smooths volatility.
- SIPs automate and reduce emotional decision-making
- Diversified portfolios significantly lower risk.
- It also encourages financial discipline in individuals by maintaining proper utilization of the hard-earned money

Stress often emerges from uninformed decisions, not from investing itself. Investing done using proper planning ensures clarity, reduces uncertainty, and promotes steady financial growth

Conclusion

Investing is indispensable for achieving financial independence, beating inflation, and reducing tax burdens. While bank deposits offer safety, they cannot match the combined benefits of wealth creation, inflation protection, and tax efficiency that diversified investments provide.

3. How can one be an effective investor?

Introduction

Investment is both a financial and behavioural journey. An effective investor would not necessarily be someone who buys and sells stocks randomly but one who makes informed decisions based on historical data, market information, and disciplined strategies. This report walks through how a beginner or intermediate investor can become effective through substantiation with data-driven reasoning, regular investment practices, and an understanding of how stock markets work.

3.1 Using Historical Data and Information for Effective Investing

Historical market data provides insights into long-term trends, patterns of volatility, influences from the macroeconomic environment, and company-specific events. By studying the historic price movements, changes in volume, quarterly results, and cycles of the industry, an investor is able to understand whether the stock demonstrates consistent growth, high volatility, or cyclical behaviour.

Large-cap stocks, represented by companies forming the Nifty 50, usually exhibit more stability than their small-cap brethren. Long-term charts give investors insight into patterns of bull runs, corrections, and phases of recovery. This comes along with investigations into the fundamentals of a company, including its EPS growth, P/E ratio, leverage, dividend track record, and quality of the management leading it.

Thus, combining historical price analysis with company fundamentals leads to more informed and effective investing.

3.2 Frequency of Investing: Daily, Monthly, or Lump-Sum?

How often to invest is a matter of the investor's financial situation and risk tolerance. Common strategies include:

a. Daily or Frequent Small Investments

The investment of a fixed sum daily is rarely done because:

- It increases transaction costs.
- Markets rarely change dramatically in just one day.
- It requires more work and attention.
- Thus, daily investments are not efficient.

b. Systematic Monthly Investments - SIP Strategy

This is the most effective method to invest for the majority of investors. Fixed-date-of-the-month investing helps in:

- Rupee-cost averaging means buying more units when prices are low and fewer units when prices are high.
- Building long-term discipline,
- Decrease emotional decision-making.
- It works well in equity markets and mutual funds because it smoothes out the volatility.

c. Lump-Sum Investing

Suitable for:

- Investors who have a sizable amount available,
- When markets are undervalued-for example, during major corrections.

However, it carries timing risk-if you invest just before a fall, you may face immediate losses. Thus, lump-sum is recommended only when market valuation is clearly favourable.

3.3 Is the Stock Market Operational Daily? What Time Should an Investor Invest?

The Indian stock market is not open 24 hours. The usual NSE trading hours are:

NSE Market Timings

- Pre-opening session: 9:00 AM – 9:15 AM
- Normal (continuous) trading session: 9:15 AM - 3:30 PM

- Post-market sessions: After 3:30 PM for order matching only (not trading)

Markets close on weekends and declared holidays.

When should you invest during the day?

- Avoid first 15–30 minutes, 9:15–9:45 AM; that is a really volatile time because of overnight news.
- Generally, the noontime period (between 11 am to 2 pm) is considered more stable and therefore suitable for long-term investors.
- The end of the day, from 3 PM to 3:30 PM, may be volatile because of institutional activity.

For long-term investors, the impact of timing during the day has little relevance compared to identifying fundamentally strong stocks.

3.4 Which Industry Should One Invest In?

Diversification is vital. Instead of investing all money in only one sector, such as IT or Telecom, it is highly recommended to spread investments across:

- IT and Technology
- Banking and Finance
- Pharmaceuticals and Healthcare
- FMCG
- Automobile
- Telecom
- Energy and Infrastructure

Diversification across industries reduces the risk of sector-specific shocks, such as changes in policy, fluctuations in global demand, or changes in technology.

3.5 Domestic vs. International Investing

Domestic investing is easier because:

- Investors have a better understanding of the Indian economy.
- Taxes and regulations are straightforward,
- Transaction costs are lower.

International investing, however, offers:

- Global diversification,
- Exposure to world-leading companies, such as Apple and Microsoft.
- Protection against domestic market downturns.

Starting domestically and gradually adding international exposure through ETFs or mutual funds is still recommended for beginners.

4.Preprocessing & EDA

As a part of cleaning/preprocessing our dataset, which in this case is the daily trade details of the Tata Consultancy Services (from 01.04.2015 to 31.03.2025), we analysed and found the days which were holidays, i.e., the days on which the National Stock Exchange did not trade any stocks. Since any time, series plot requires time instances to be equally placed, these holidays create inconsistencies/discontinuities which break the equally spaced intervals. Therefore**, we encoded the dates using indices**, so that time lags can be computed meaningfully.

Next as a part of our Exploratory Data Analysis (EDA) we visually inspected if there is any trend or seasonality in the series. We also checked if the given series is stationary or not (i.e. if it has a constant mean or not), using the **ADF (Augmented Dickey Fuller) test**.

ADF test for stationarity:

The ADF test checks if the given time series has a unit root or not (i.e. whether time series wanders about a simple pattern or not). The presence of a unit root indicates non-stationarity of the series.

The ADF test equation:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \sum \delta_i (\Delta y_{t-i}) + \epsilon_t \quad \text{(summation over the entire lag)}$$

Where:

- Δy_t= first difference of the series
- α= intercept
- βt = optional trend
- γ= coefficient of y_{t-1}
- p= lag length to remove autocorrelation

Next, test hypothesis on γ :

H_0: γ=0 (unit root exists → series is non-stationary)

H_1: γ<0 (no unit root → series is stationary)

Therefore, to reject H_0, we calculate the value of the test-statistic and check if it lies in the critical region corresponding to level of significance = 0.05 (corresponding critical value for the statistic = -2.86).

```
=========== ADF TEST ===========
ADF Statistic : -1.8385927328068246
p-value       : 0.361476696015011
Critical Values:
   1% : -3.4330306663366255
   5% : -2.862724291785039
   10% : -2.56740044647283

→ ADF says: NOT stationary (fail to reject H0)
```

Fig 2 : ADF test results

As we can see that the p-value > 0.05, we fail to reject the null hypothesis, which means there is strong evidence that the series is non-stationary.

Fig 3 : Best fit line for trend along with 30 day and 90 day rolling average

As we can see from the trend analysis diagram above, the closing price seems to have a net positive trend over the 10 years frame as indicated by the trend line. However, the dip around the region mid-2018 to early 2020 could have occurred due to the issuance of bonus shares by TCS on a 1:1 basis which doubled the number of outstanding shares in the market while the market value was constant leading to an adjust in prices , in addition to this the COVID 19 pandemic could have been another reason for the dip (note that these add to the cyclic component of the time series), but the market seems to have recovered post that.

Next, to analyse the presence of seasonality in the data, we plotted the **ACF (Auto-correlation Function plot).**As we can see from the plot below , that the evidence for seasonality is not very strong (because the spikes are not repeating and very high for almost all lags , however we can't conclude about the total absence of seasonality from this).We arrived at the same conclusion from the **multiplicative** season-trend decomposition graph too (the amplitudes of the seasonal waves are too small).It must be noted that it is better to follow a multiplicative decomposition of the stock price data because stock prices show fluctuations which are proportional to the level of the series which typically defines a multiplicative relationship. In our case we first found out the STL decomposition of the data and then used a logarithmic transformation to obtain the multiplicative decomposition.

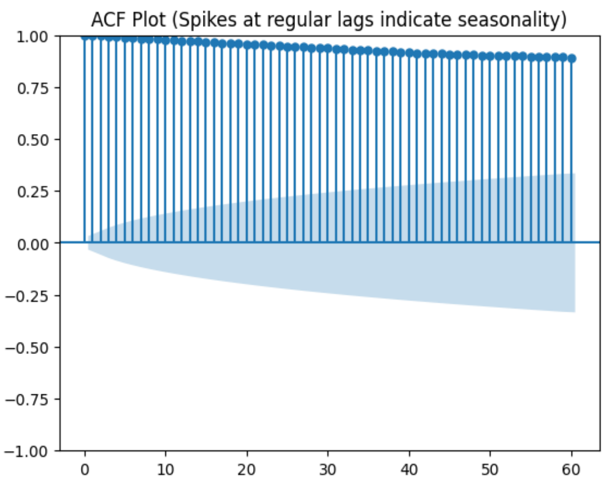
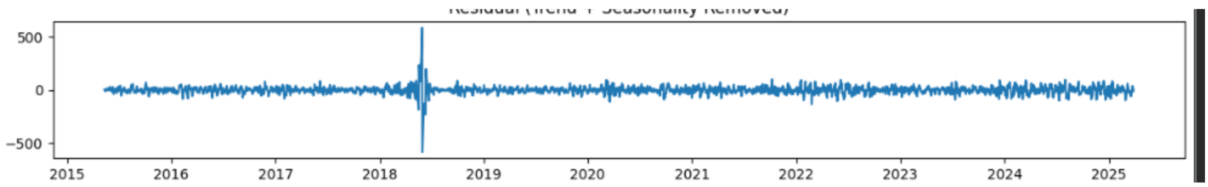


Fig 4 : ACF Plot

Fig 5 : Multiplicative Decomposition

4.1 Underlying Probability Distribution

To check if the given time series follows any underlying follows some probability distribution, we remove the trend and the seasonality from the data, i.e., we first make it stationary (this is because any probability distribution is defined for a random variable whose statistical properties like mean do not change over time).Therefore , it is necessary to note that we carry out the distribution analysis only on the residuals which would be truly random.(Given below in **Fig 6** is the residual plot for the time series)



Therefore, once the series has been made stationary, we try to fit multiple distributions through the data. We first perform the **Shapiro-Wilk test** to detect normality within the data. However, we obtained a p-value = 1.2481024858172004 x 10^{-57} < 0.05 which means the residuals do not follow normal distribution.

Since the series fails to follow a norma distribution, we perform the **Kolmogorov–Smirnov** (K–S) test which is used for testing the possibility of the data coming from different distributions. In K-S test, we compare the empirical CDF of the data with the theoretical CDF of the target distribution, if the difference is small, then the data most likely fits that particular distribution. Therefore, lower the KS statistic value, better the approximation from the respective distribution.

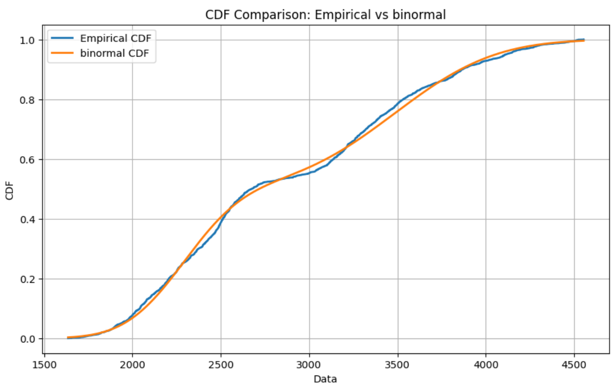
	distribution	p_value	ks_stat
9	binormal	1.232199e-02	0.032275
4	weibull_min	8.359718e-15	0.082191
3	gamma	3.998097e-22	0.100945
1	lognorm	2.205712e-22	0.101543
6	t	5.724874e-33	0.123566
0	norm	5.724783e-33	0.123566
7	cauchy	2.066306e-49	0.151486
8	laplace	9.471678e-65	0.173459
2	expon	6.514435e-69	0.178946
5	weibull_max	0.000000e+00	0.812896

```
Binormal parameters:
Weight (w):      0.5034292861755901
Mean 1 (mu1):    2288.959291379935
Sigma 1:         261.18201864591185
Mean 2 (mu2):    3481.9374296248307
Sigma 2:         449.0044049512353
```

Fig 7: KS statistic and p-values for various distributions,

Fig 8 : Binormal Dist. parameters

As we can see above, the best fitting distribution for the residuals is a **Bi-Normal** distribution (with parameters given in Fig 8). Note that these parameters are the MLE (Maximum Likelihood Estimators) Estimate for the original parameters and they are biased (because there are no closed form representations of these estimates).



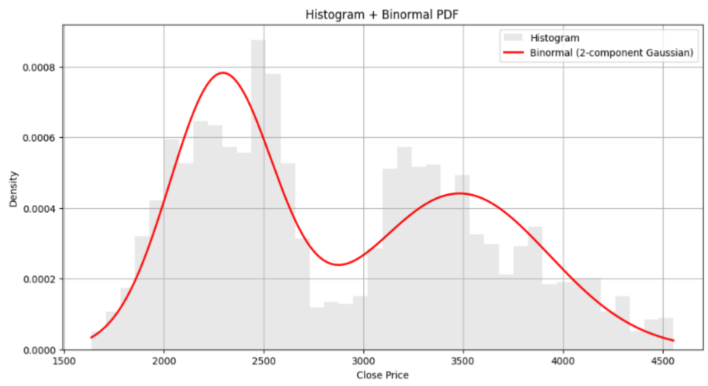


Fig 9 : CDF plots approximation vs actual

Fig 10 : PDF fit on the original data histogram

5.Statistical Forecasting Models

We had split the time series data into train/test split for building the forecasting the model and then defining the goodness of fit through appropriate metrics like **RMSE** (root man squared error) , **MAE** (mean absolute error) or **MAPE** (mean absolute percentage error). We trained the following statistical forecasting models:

	Parameters
1)AR(Autoregressive)	**p**
2)MA (Moving Average)	**q**
3)ARMA (Autoregressive Moving Average)	**p,q**
4)ARIMA (Autoregressive Integrated Moving Average)	**p,d,q**
5)SARIMA (Seasonal Autoregressive Integrated Moving Average)	**p,d,q,P,D,Q,s**

Where:

P = Number of lag observations (past values) used in the model, i.e., the number of past observations the new target value is dependent on

q = Number of lagged forecast errors included in the model

d = Degree of differencing (to make series stationary)

P = number of seasonal lag observations (past values) used in the model

D = number of seasonal lagged forecast errors included in the model

Q = seasonal Degree of differencing s = length of the seasonal cycle

The basic difference between p, d, q and P, D, Q is that the latter are the seasonal analogues of the former non – seasonal parameters that are taken into account when the series has seasonal patterns.

AR (Autoregressive)

Equation:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

Where:

- X_t= current or forecast value
- $\phi_1, \ldots, \phi_p$= AR coefficients
- p= lag order
- ϵ_t= white noise at time t (c = constant term)

<u>Use Case</u>: Suitable for a stationary time series (no trend or seasonality)

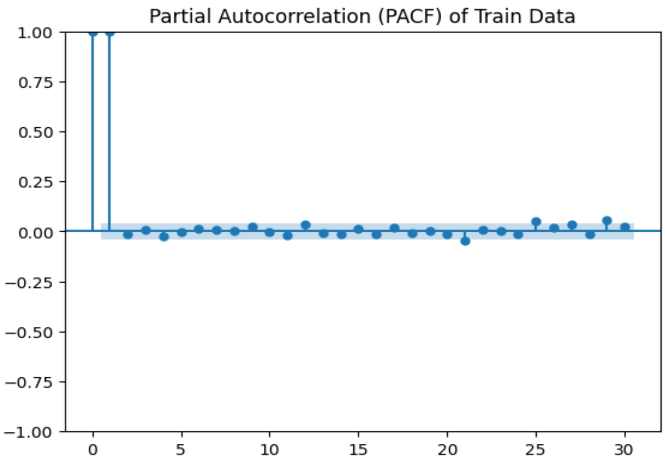
<u>Analysis and Implementation</u>: The AR model is the simplest forecasting model which captures the persistence in data efficiently. However, it fails when the series is not stationary or has sudden shocks or seasonality.

To estimate p, we plot the PACF (partial auto-correlation graph) and see the number of significant spikes. We can see that there is one significant spike in the graph which means the value of the target variable for a particular time frame is related only to its value in the previous time frame and no intermediate frames or beyond.

Therefore, we used an AR (1) model to forecast the series for the test split and obtained the following, RMSE, MAE and MAPE scores:

```
Training MSE: 3034.0832012724413
Training RMSE: 55.08251266302619
Forecast MSE : 90231.74743562123
Forecast RMSE: 300.38599740271053
Forecast MAE : 275.34481679055136
Forecast MAPE: 7.688126779121535
```

Fig 11 : Performance metrics for AR(1) model



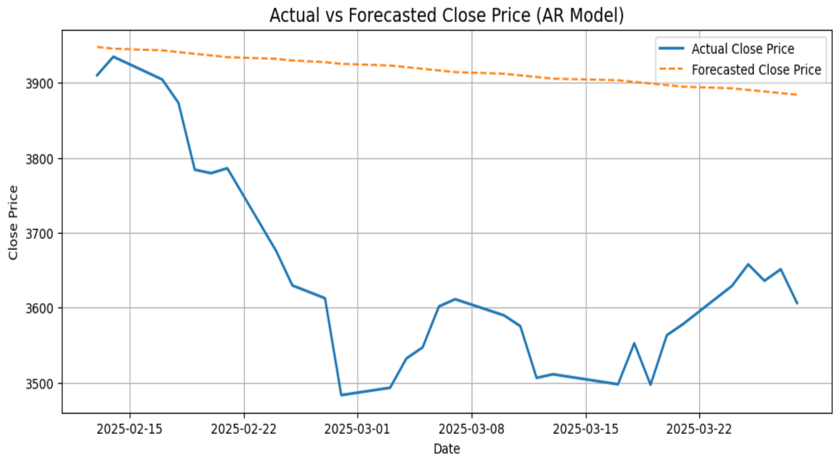


Fig 12 : PACF plot to obtain value of p

Fig 13 : Test data forecast vs actual

<u>MA (Moving Averages)</u>

Equation:

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Where:

- ϵ_t= white noise at time t
- $\theta_1, \ldots, \theta_q$= MA coefficients
- q= order of moving average

<u>Use case</u>: Suitable for stationary series where random shocks or noise influence the series or for short term modelling about the mean.

<u>Analysis and Implementation:</u> The MA model is yet another simple forecasting model which captures noise effects well. However, if used alone, it fails to model any trend that the series may have.

So, we took q or window size = 5 and built the forecasting model and got the following metrics:

```
Training MSE : 6727.102851292765
Training RMSE: 82.01891764277778

====== Forecast Error Metrics ======
Forecast MSE : 7779.582923333324
Forecast RMSE: 88.20194398840269
Forecast MAE : 74.15233333333327
Forecast MAPE: 2.0385333741202283
```

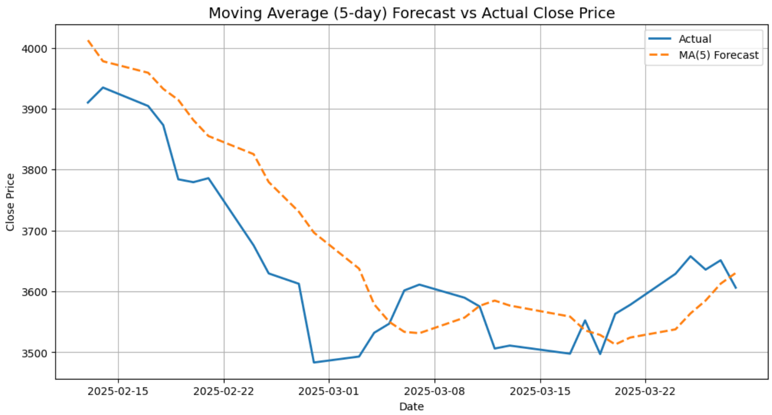


Fig 14 : MA(5) model performance

Fig 15 : Test data forecast vs actual

ARMA (Autoregressive Moving Averages)

Equation:

$$X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

Where:

- ϵ_t= white noise at time t
- $\theta_1, \ldots, \theta_q$= MA coefficients
- q= order of moving average

Use case: Can be used to model stationary time series with lag dependencies as well as noise effects

Analysis/Implementation: Note that ARMA is a special case of ARIMA of the form (p,0, q). It basically combines both AR and the MA models. The choice of p,q can be done using multiple methods. We decided to stick with the combination that minimises the **AIC** (Akaike Information Criterion). AIC tells us how good a model is while imposing a penalty for using too many parameters. A low AIC models the data well without overfitting.

For, our model, a low AIC was obtained for p=2, q=1, the corresponding performance is as follows:

```
ARMA Train RMSE: 55.503923046970556
ARMA Train MAE : 30.54817831986687
ARMA Train MAPE: 1.1056633094985415
ARMA Test RMSE: 285.5730300293015
ARMA Test MAE : 261.0955936260554
ARMA Test MAPE: 7.688126779121535
```

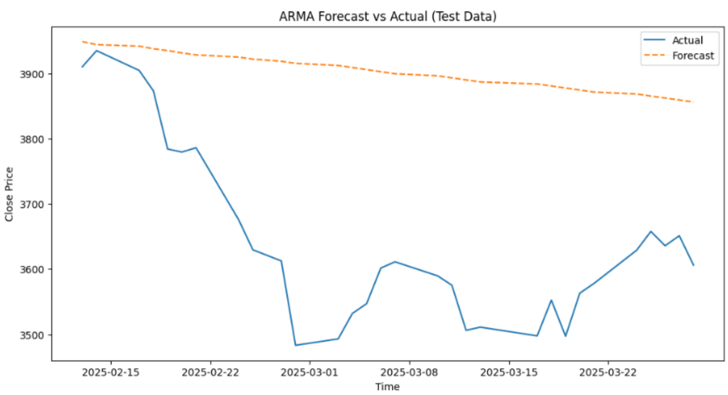


Fig 16 : ARMA (2,0,1) performance metrics	Fig 17 : test data forecast vs actual

ARIMA (Autoregressive Integrated Moving Averages)

Equation:

$$\Delta^d X_t = c + \phi_1 \Delta^d X_{t-1} + \cdots + \phi_p \Delta^d X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

Where:

- ϵ_t= white noise at time t
- $\theta_1, \dots, \theta_q$= MA coefficients
- q= order of moving average
- $\Delta^d X_t = (1 - B)^d X_t$ is **d**$^{\text{th}}$ differenced series to make it stationary.

<u>Use case:</u> Can be used to model when the series is non stationary but with no clear seasonality (similar to our case).

<u>Analysis/Implementation</u>: With ARIMA, we can handle any non-stationary series using the differencing which removes non-stationarity.

As stated earlier, there are 3 parameters in ARIMA -p,d,q. The choice of d is made such that after differencing with that factor the series becomes more or less stationary. We can use the ADF test again and again till it stops failing and obtain the final best value of d for our model. Once the series has been made stationary after differencing with degree d, we can obtain a combination of p,q using the AIC in a similar way we did for ARMA and proceed to build our forecasting model.

For our model, the best parameters we derived were: p=0, d=2, q=0 and the performance was as follows:

```
ARIMA Train RMSE: 96.49598741108176
ARIMA Train MAE : 46.3196136103406
ARIMA Train MAPE: 1.686954335212769
ARIMA Test RMSE: 147.7182179579341
ARIMA Test MAE : 123.26833333333585
ARIMA Test MAPE: 3.4451573842630494
```

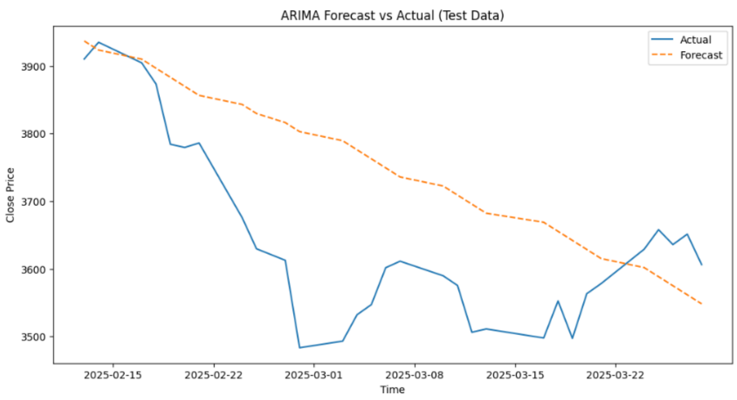


Fig 18 : ARIMA(0,2,0) performance metrics ; Fig 19 : test data forecast vs actual

SARIMA (Seasonal Autoregressive Integrated Moving Averages)

Equation:

$$\Phi_P(B^s)\phi_p(B)\Delta_s^D\Delta^d X_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t$$

Where:

- $\phi_p(B)$, $\theta_q(B)$= non-seasonal AR and MA
- $\Phi_P(B^s)$, $\Theta_Q(B^s)$= seasonal AR and MA
- d= non-seasonal differences, D= seasonal differences
- s= seasonal period

Use Case: Can be used to model a series with both trend and seasonality.

Analysis/Implementation: The SARIMA forecasting model can become quite a complex model with so many parameters; however, it is very flexible for seasonal forecasting. A good combination of the parameters usually requires domain knowledge; p,q,d can be estimated exactly like we did in ARIMA, but only after de-seasonalising the data.

For our series, we got the best results with the following combination:

Non-seasonal (p,d,q) = (0, 1, 1)

Seasonal (P, D, Q, s) = (0, 1, 1, 5)

```
====== SARIMA TRAIN ERRORS ======
SARIMA Train RMSE : 82.933
SARIMA Train MAE  : 32.576
SARIMA Train MAPE : 1.186%

====== SARIMA TEST ERRORS ======
SARIMA Test RMSE  : 347.596
SARIMA Test MAE   : 320.418
SARIMA Test MAPE  : 8.942%
```

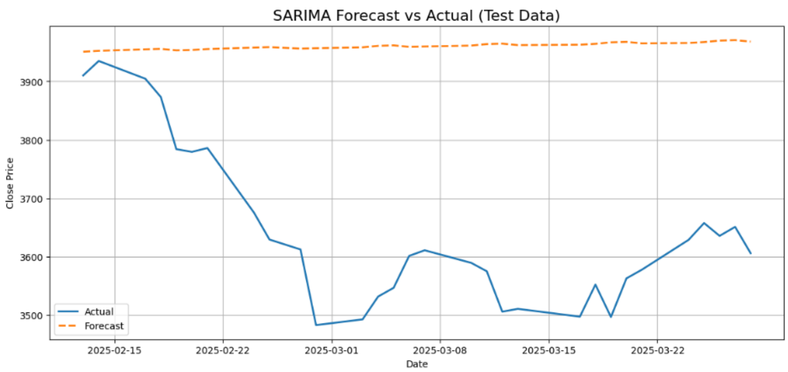


Fig 20 : SARIMA performance metrics Fig 21 : Test forecast vs actual

Therefore, the SARIMA model performs slightly worse than the other models may be because we are trying to force some seasonality onto our data when there is clearly no actual seasonality present.

LSTM (Long Short-Term Memory)

Additionally, we also trained a neural-network based machine learning model called Long Short-Term Memory which is often used in our dataset is time-series based. We used an LSTM because they learn long term dependencies better than RNNs (Recurrent Neural Networks). LSTM comes in handy when there are non-linear regime shifts in the data that models like ARIMA/SARIMA cannot model well enough.

Neural Networks are considered data-hungry approaches which is why we should ideally consider data that goes back more than 10 years; however, we still managed to get good accuracy with the available time frame. As we can see below, the accuracy for the LSTM model was better than the best statistical model we could build with just 1.27% error.

```
LSTM Train RMSE: 63.85532567540175
LSTM Train MAE : 41.24218110316875
LSTM Train MAPE: 1.443747600634919
LSTM Test RMSE: 53.960789452129795
LSTM Test MAE : 45.94456054687498
LSTM Test MAPE: 1.2681869538707171
```

Fig 22 : LSTM performance metrics

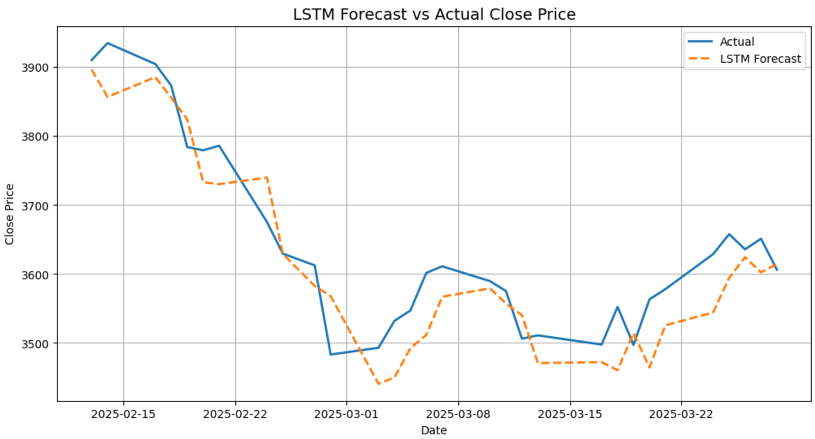


Fig 23 : Test forecast vs actual

Summary

Below is the final summary of all the statistical forecasting models we used and their test MAPE:

Forecasting model	MAPE
AR(1)	**7.688%**
MA(5)	**2.0385%**
ARMA(2,0,1)	**7.0688%**
ARIMA(0,2,0)	**3.44%**
SARIMA(0,1,1)(0,1,1,5)	**8.942%**

Conclusion

It requires discipline, efficient knowledge of the market, diversification, and data-driven decisions while investing. While analysing historical trends, adopting systematic month-on-month investments, knowing market hours, diversification across industries, and balancing domestic and international exposure, an investor can build long-term strong wealth efficiently with risk management.

REFERENCES

- Elements of Forecasting Book by Francis Diebold
- Geeks For Geeks Time Series Analysis and Visualisation in Python
- Time Series Analysis with Python
- Time series – tutorial for beginners
- NSE – India for Dataset
- Nifty 50 Documentation
- Meshram, A. S. (2022). *An analysis of banking saving rates in India*. International Journal for Multidisciplinary Research (IJFMR), 4(3), 66-68. Retrieved from https://www.ijfmr.com/papers/2022/3/25015.pdf
- National Savings Institute. (n.d.). https://www.nsiindia.gov.in/(S(cbazf4nq1i2v0sy25w4sajag))/InternalPage.aspx?Id_Pk=132
- BMS Money. (2025, November 1). *Decades of NIFTY 50 performance: A comprehensive analysis of returns from 1991 to 2024*. Retrieved from https://www.bmsmoney.com/article/full/decades-of-nifty-50-performance-a-comprehensive-analysis-of-returns-from-1991-to-2024/
- ET Money. (n.d.). *Best aggressive hybrid mutual funds to invest in India 2025*. Retrieved from https://www.etmoney.com/mutual-funds/hybrid/aggressive-hybrid/68
- Poorna Pal. (2021, October 3). *India's inflation and interest rates through these past 10 years*. PGurus. Retrieved from https://www.pgurus.com/indias-inflation-and-interest-rates-through-these-past-10-years/

GITHUB repository for our code: https://github.com/GeneraLIroh04/MATH-F432-Assignment-Closing-Price-Prediction-Model-

___END___